Pay Enough - Or Don't Pay at All

Gneezy, U.; Rustichini, A.

Publication date:
1998

Pay Enough-Or Don't Pay At All

Uri Gneezy** and Aldo Rustichini***

May 1998

Abstract: Economics seems largely based on the assumption that monetary incentives improve performance. By contrast, a large literature in psychology, including a rich tradition of experimental work, claims just the opposite. In this paper we present and discuss a set of experiments designed to test the effect of different monetary compensations on performance.

In our experiments we find that whenever money is offered, a larger amount yields a higher performance. It is *not* true, however, that offering money always induces a higher performance: participants who were offered a small payoff gave a worse performance than those who were offered no compensation at all. These results suggest that the behavior of participants is influenced by their perception of the contract that is offered to them. When the contract offers money the environment is perceived as monetary, and participants respond in a qualitatively different way in monetary and non-monetary environments.

In a different set of experiments we test subjects who, acting as principals, have to provide the appropriate incentive to agents. We show that principals do not anticipate the drastic difference in behavior. The vast majority of principals seem to think incorrectly that a larger compensation is unambiguously a better incentive.

Keywords: Monetary incentives, performance, motivation, principal-agent.

JEL Codes: D1, D8, D9.

* We wish to thank Doug Dejong, Martin Dufwenberg, Georg Kircheisgter, and Ariel Rubinstein for helpful comments and suggestions. Special thanks go to Eric van Damme for his comments as well as for his help in raising the money needed for the experiments.

** University of Haifa, Israel. Email: gneezy@econ.haifa.ac.il

*** CentER for Economic Research, Tilburg University, The Netherlands. Email: aldo@kub.nl

1. INTRODUCTION

A major problem facing most principals is how to set the incentives for their workers right in order to improve performance. A great deal of discussion is devoted to this question in two different fields: economics and psychology. The literature in the two fields seems to make opposite statements on the effect of monetary incentives on performance. Economics is largely based on the assumption that monetary incentives improve performance.[1] On the other hand a large body of literature in psychology, including a rich tradition of experimental work, claim just the opposite. In this paper we present and discuss a set of experiments designed to test specifically the effect of monetary compensation on performance. On the basis of our results we argue that both claims are, for different reasons, correct.

In the results of our experiments the performance, and so presumably the effort, of the agents is *not* monotonically increasing in the amount of monetary compensation offered to them, even when the amount paid is increasing with the performance. This observation seems to confirm the claim in the psychology literature that money may have detrimental effects on motivation. On the other hand we observe that *conditionally* on the fact that money is offered, more money *does* induce a larger effort.

It seems important to determine the reason for these seemingly contradictory observations. Here we anticipate a tentative explanation, based on our results. An experimental environment (just as, presumably, a real life environment) defines a contract for the individuals. Our results suggest that the critical element in determining their

[1] There are, however, interesting exceptions to this rule, some of which are recent: for instance see the work of Titmuss (1970), Arrow (1972), Frey (1994), Frey, Oberholzer and Eichenberger, (1996), and Frey and Oberholzer-Gee, (1997).

behavior is their perception of the contract they are facing. If we compare situations in which a monetary compensation is offered, then we see that a higher compensation induces a higher effort.[2] On the other hand, a contract that does not mention contingent payoffs may induce an effort higher than the one induced by a contract offering only a modest amount of contingent payoff. In the first case subjects perceive the contract as defining a monetary environment, and choose the effort comparing the monetary reward with the disutility produced by the effort.

This explanation is in the spirit of the "fair wage" hypothesis. Akerlof (1982) and Akerlof and Yellen (1988, 1990) argue that fairness-oriented behavior of workers may lead to involuntary unemployment. The main assumption they make is that a positive relationship between work effort and wages exist. Given this relationship, It may be profitable for employers to pay more than the market clearing wages.

2. RELATED LITERATURE

The evidence we described and the tentative explanation we have offered differ in an important way from those offered in the psychology literature. So, before we proceed to present and discuss our experiments and results, we review some of the evidence and analysis in that literature.

The ''hidden cost of reward''.

---

[2] We would like to note that we mean payoffs which are contingent on performance, and not payoffs which are given as show-up fee.

The theme of the "hidden cost of reward" appears explicitly in the psychology literature in the early seventies: it is also the title of a famous book collecting essays in the topic (Lepper and Greene, 1978). This line of research began as reaction of the new trend of cognitive psychology to the dominant behaviorist paradigm. To understand the reasons for this new development it is useful to recall that, according to the classical conditioning, an activity in itself neutral or even mildly unpleasant but positively rewarded may eventually acquire the positive valence of the reward. In this case, past reward has in the long run a positive effect on the performance of that activity. In fact from a strictly behaviorist point of view motivation in itself has no meaning: the only scientifically relevant (and observable) aspect are behavior, and the relation between reward and behavior. The main prediction of the behaviorist theory is that the performance of some particular task is increased if in the past it has been repeatedly associated with positive rewards.

Intrinsic versus extrinsic motivation.

The cognitive approach offered an alternative view: an activity has a motivation of its own, independent of any reward, which was called *intrinsic motivation*. An operational definition of this concept is easy: we may say that an activity is intrinsically motivated if it is performed when there is no external reward associated with it. A conceptual definition is not that easy. For instance, Deci (1975) defines intrinsically motivated behaviors, as those in which a person engages to feel competent and self-determining. A full discussion of the reasons for such a definition would take us far from our purposes, so we refer the reader to the book of Deci (1975).

The focus of the literature became soon the question: do rewards (or any extrinsic motivation) reduce intrinsic motivation? The importance of this question is clear: a "hidden cost of reward" (that is, a reduction in the intrinsic motivation following a reward) was a powerful way for cognitive psychologists to show that the basic assumption of behaviorism was flawed. A rather large set of experiments showed that a reduction was indeed taking place.

Experimental evidence in the literature

As an illustration we may consider the classic experiment reported in Deci (1971). He had college students play with a puzzle (called "Soma"), in three succeeding sessions. In the first session participants were left to play freely with a puzzle. In the second session, one group received payment conditional on the solution of the puzzle and the control group did not. In a third session the subjects were then left to play freely with the puzzles. The amount of time spent in the free activity performed in the first and third session was taken as a measure of intrinsic motivation. Deci found that the experimental group spent less time playing with the puzzle in the third session than the control group, and concluded that the reward offered had decreased the intrinsic motivation of subjects in the first group.

In the following years a rich set of experiments was conducted to further test the effect of rewards on intrinsic motivation. It is important to note that these experiments were trying to test the effect of past reward on future motivation. (Our experiment, on the other hand, will focus on the effect of a reward on present performance.) The main findings of these experiments were:

(i) Positive rewards, in particular monetary rewards, have a negative effect on intrinsic motivation. If a person is rewarded for doing an interesting activity his intrinsic motivation decreases. But also the converse was true: if a person is insufficiently rewarded for performing a dull activity his intrinsic motivation increases.

(ii) The negative effect is significant only if the reward is contingent on the performance; the subjects who are paid a fixed positive amount, independent of their performance do not display reduction in intrinsic motivation. The same happens if subjects are informed of the fact that a reward will be paid only *after* they have performed the activity (so that they were not expecting the reward at the moment of performing it).

(iii) A positive feedback, provided by sentences like: "Very good, that's the fastest this one has been done yet", has positive effects on intrinsic motivation. Conversely a negative feedback ("Well, most people were able to solve this one, but let's go on to the next") has a negative effect (Deci, Cascio and Krusell, 1973).

On two issues this literature was silent. First, there is no attempt to determine precisely if the motivation is task-specific or environment-specific. For instance, in the experiments reported in Deci (1971) that we have already mentioned one might wonder if the subjects, left alone with a completely different type of puzzle, would still display the reduction of intrinsic motivation. This might be observed if the subjects that had been promised a payment in the first session displayed less activity on the new puzzle, compared to the control group. If they did, one might conjecture that the subjects were originally

motivated to produce effort in the situation created by the experiment, irrespective of the specific task: this would be evidence of an environment-specific motivation. Suppose on the other hand that they did not: then it would be natural to conjecture that the motivation was attached to that specific task (the "Soma" puzzle, in this instance). In fact most of this literature seems to assume implicitly that the motivation is task specific. The distinction is important if one wants to understand precisely how wide in scope and persistent in time are the effects of rewards on motivation and performance.

Second, very little attention was given to the performance, rather than motivation, of the subjects in the different treatments. The work by Kruglansky, Freedman, and Zeevi (1971) is a partial exception: a group of teenagers was asked to perform several tasks, and some of them were promised a reward (presumably positive) consisting of a guided tour of the psychology department of a nearby university. The performance of the rewarded students was lower.

The theory: Self-perception and attribution theories.

A brief sketch of the theory provided as an explanation of this evidence may be helpful to complete the picture. A widely accepted explanation was based on the cognitive approach. According to this approach a subject may use external evidence to find a reason for his own actions. In a process called later *attribution* a subject associates his own observed behavior to a motivation. The original idea of attribution goes back to Heider (1958). According to Heider, a person constantly tries to understand the world as something non-random. In this process, he tries to link every event, including his own behavior, to the conditions that underlie it.

A development of the original idea of Heider is in Bem's Self-Perception Theory (see Bem 1965 and 1967). The main idea is that subjects, in order to explain their own behavior, use the evidence they have as any outside observer would do. In particular if subjects see that they do an activity when they have been offered a monetary incentive, they explain their motivation as provided by that monetary incentive. Deci formulated the application of this idea to the analysis of the effect of rewards as the "overjustification hypothesis" (Deci 1971): when some external factor (like some compensation) is added, it is likely that the subjects will attribute the motivation to that factor. So when offered, a reward replaces the intrinsic motivation which then becomes less important. When the extrinsic motivation is removed, this reduction in the intrinsic motivation produces a reduced effort.

All the findings we have briefly reported are consistent with this cognitive evaluation theory. For instance the evidence that larger rewards have larger and negative effects may be explained as a consequence of the discounting principle, (see Kelley 1967 and 1971). The principle states that when behavior occurs in presence of several multiple causes, then the attribution of the behavior to these different causes is less than it would be if only one or fewer of these causes were present.

Another explanation was provided along the lines of Festinger's cognitive dissonance theory (see Festinger, 1957). A subject who is doing a dull task may, to reduce the dissonance created by observing himself doing something uninteresting, convince himself that the task is attractive, and therefore increase the intrinsic motivation.

The line of research we have briefly discussed addresses an extremely ambitious question, which is in fact the same as the behaviorist school was addressing: "How do rewards affect behavior in the long-run?" If we use concepts and definitions more familiar to economists or decision theorists, the question may be formulated as "How are preferences over consequences affected by past experiences?" For most of this literature the object of study is the long-run effect of rewards. It is important to note that the rewards were very rarely repeated in the experiments we described. This attracted a criticism from the behaviorist side of the experimental evidence provided by the cognitive school on effects of rewards on motivation (see for instance Reiss and Sushinsky (1975a and 1975b)). They pointed out that a proper analysis of these effects on the long run would require an extended series of reinforcements rather than single-trial reinforcement.

Experimental economics: the Dominance condition.

The issue of the effect of monetary compensation on the behavior of subjects marks of course a major distinction between the experimental methods in psychology and economics. For economists, a monetary payoff is a very important requirement, which is needed to give to the experimenter the control over the environment of the experiment. The monetary payoff offered to the subject must be sufficient to overcome the subjective factors, which are difficult to evaluate and extraneous to the theory. This is the *dominance* condition (see Smith (1982)). A quantitative condition for dominance is that the payoff must be large enough to insure that the reward corresponding to the experimental hypothesis are "perceptibly and motivationally greater" than the rewards corresponding to the alternative hypothesis. A *qualitative* condition is, however, that the subjective factors

are independent of the monetary payoffs. Our results suggest that this assumption of independence between monetary payoffs and subjective factors may typically not be satisfied. In particular, this would put in doubt the extension of some results, obtained with zero monetary payoffs, to situations where the payoffs are strictly positive.

3. EXPERIMENTAL DESIGN AND PROCEDURE

In this paper we address the issue of the effect of incentives on performance, rather than motivation. In addition we set up this test in an environment which is quantitatively precise, and perhaps more precisely defined.

Our experiments were designed to test two basic assumptions, which seem widely accepted among economists. First, in standard economic theory preferences over outcomes are assumed to be independent of the environment in which the person operates, and in particular from the contract offered. To illustrate this assumption we may consider our IQ test experiments: subjects were asked to answer questions out of an IQ questionnaire, and were then informed about their performance, and are paid money or not accordingly. The relevant outcomes for the subject here are the effort and time, the possible entertainment value of answering the questions, the opinion of the people who will know his results, and so on. Second, a basic assumption of an economist in this situation would probably be that, since money is a good, if two contracts only differ in the amount of money paid for an observed performance, the contract offering a higher amount of money conditional on a better performance should induce a higher effort and as a consequence a higher performance.

The first set of experiments involves "agents", and was designed to test these two assumptions. To anticipate our main results, we may say that the second assumption will be verified, while the first will not.

A second issue we address is the perception that people have of the effect of monetary compensation on performance. The specific question is motivated by the results observed in the first set of experiments, which seem to indicate clearly that the offer of a small compensation may induce a worse performance than the complete absence of monetary compensation. Do people anticipate this? The second set of experiments, involving "principals" is designed to test whether they do. As we shall see, the answer of our experiments is that they do not.

We conducted two different sets of experiments, called the "IQ experiment" and the "donation experiment", for reasons that will soon be clear.

The IQ experiment.

The experiment was conducted at the University of Haifa with 160 students participating in the role of "agents", and 53 in the role of "principals". The participants were male and female undergraduate students from all fields of study, with an average age of 23. We start by describing the procedure for the agents.

At the beginning of the experiment each student in the role of the agent received an introduction, in which (s)he was told that the experiment would take 45 minutes. They were also told that they would be asked to answer a quiz consisting of 50 problems taken out of a psychometric test used to scan candidates to the university. This is a test similar to the GMAT exam: the participants were told that this is a sort of IQ test. The problems

in the quiz where chosen to make the probability of a correct answer depend mostly on effort. In particular, an emphasis was given to questions involving reasoning and computation rather than general knowledge questions. Each student was then informed that he would receive 60NIS[3] for showing up to the experiment. The instructions are presented in the Appendix.

We had four different treatments, according to the payment for correct answers. The payoffs for each correct answer were 0, 0.1, 1, and 3 NIS in treatment 1, 2, 3, and 4 respectively. In each treatment 40 different students participated. After the introduction, the quiz was distributed. Participants were not allowed to have any material on their tables except the quiz itself, and were told that only those who would stay until the end of the experiment would be paid. No clarifying questions by students were allowed during this time. At the end of the experiment participants were told where and when to come and pick their earnings.

The 53 students in the role of principals were given a short introduction. In this introduction they were explained the task that the "agents" participating in treatment 1 and 2 will be given. The principals were also told that they would be paid 1 NIS for every correct answer that the agent with whom they would be matched would answer in the IQ test. Then the principals had to choose whether to pay 10 cents or nothing to the agent for every correct answer (out of the 1NIS they get). They were told that the other person will know in advance how much he is going to be paid for every correct answer, but that he will not know that the principal had to chose first whether to pay them 0 or 10 cents. So the agent would not know that the principal could choose how much to pay him. This was

[3] At the time of the experiment, 1 NIS was worth approximately $2.

the only decision the principals had to make. At the end of the experiment participants were told where and when to come and pick their earnings.

The donation experiment.

In Israel there are a few "donation days" every year. Each of these days is devoted to a society, which needs to collect donations from the public for some purpose, like cancer research, invalid children, etc. In each of these days, high-school students go from door to door and collect money for that purpose. Normally, the students are organized in groups according to the class in which they study. The class is divided into couples of students. Each couple receives coupons, which serves as receipts to the people who donate money. In the donation mentioned in this experiment each couple received coupons with the total amount of 500 NIS.

The amount collected by each couple in the donation day depends mostly on the effort invested: The more houses they visit, the more money they collect. This is especially true because the students do not have to "sell" the donation, since most people are already familiar with it. We had 90 high-school students around the age of 16, participating in three experimental treatments. Each treatment was conducted in two groups of participants, each with 15 couples. In the discussion that follows we report the two groups of each treatment together.

In **treatment 1a**, an experimenter appeared in front of each of the groups and told them about the importance of the donation they will collect, and that the society is interested in motivating them to collect as much money as possible. They were told that

the results of the collection would be published, such that it will be public knowledge how much money was collected by each couple.

Treatment 1b was done in a similar way, a part from the fact that each couple was promised 1% of the amount that the two of them would collect. In **treatment 1c** each couple was promised 10% of the amount they will collect.

In the donation experiment we also had a group of students that played the role of "**principals**". These students were told that they would be matched randomly with one couple who already collected money, and will be paid 5% of what this couple collected. The choice that the principals were asked to make was from which group they want us to choose the couple: from the group that did not receive any payoff (treatment 1a) or from the group of students who received 1% of what they have collected (treatment 1b). In other words, we asked them to guess which treatment yields better realizations.

4. RESULTS

In this section we report the results of the experiment and some basic statistics.

The IQ experiment: the agents.

The number of correct answers given by each participant is presented in Table 1.

Obs.#	No Payment (treatment 1)	Obs.#	10 cents (treatment 2)	Obs.#	1 NIS (treatment 3)	Obs.#	3 NIS (treatment 4)
1	**49**	41	**50**	81	**49**	121	**50**
2	**48**	42	**44**	82	**47**	122	**50**
3	**48**	43	**44**	83	**47**	123	**47**
4	**45**	44	**43**	84	**46**	124	**45**
5	**42**	45	**40**	85	**46**	125	**44**
6	**42**	46	**39**	86	**45**	126	**44**
7	**42**	47	**36**	87	**44**	127	**44**
8	**40**	48	**35**	88	**44**	128	**43**
9	**37**	49	**35**	89	**44**	129	**42**
10	**37**	50	**35**	90	**43**	130	**41**
11	**37**	51	**34**	91	**41**	131	**41**

12	**37**	52	**34**	92	**41**	132	**39**
13	**36**	53	**32**	93	**41**	133	**39**
14	**36**	54	**32**	94	**40**	134	**39**
15	**36**	55	**31**	95	**40**	135	**38**
16	**35**	56	**30**	96	**38**	136	**38**
17	**34**	57	**26**	97	**38**	137	**37**
18	**34**	58	**26**	98	**38**	138	**37**
19	**34**	59	**26**	99	**38**	139	**37**
20	**31**	60	**26**	100	**37**	140	**37**
21	**31**	61	**24**	101	**34**	141	**37**
22	**31**	62	**23**	102	**33**	142	**36**
23	**31**	63	**23**	103	**33**	143	**36**
24	**29**	64	**22**	104	**33**	144	**34**
25	**29**	65	**21**	105	**31**	145	**33**
26	**24**	66	**21**	106	**31**	146	**31**
27	**23**	67	**21**	107	**30**	147	**31**
28	**23**	68	**19**	108	**29**	148	**31**
29	**23**	69	**19**	109	**29**	149	**28**
30	**22**	70	**13**	110	**29**	150	**27**
31	**22**	71	**11**	111	**28**	151	**26**
32	**20**	72	**8**	112	**28**	152	**25**
33	**20**	73	**0**	113	**26**	153	**25**
34	**18**	74	**0**	114	**23**	154	**21**
35	**7**	75	**0**	115	**22**	155	**20**
36	**3**	76	**0**	116	**22**	156	**20**
37	**0**	77	**0**	117	**22**	157	**19**
38	**0**	78	**0**	118	**21**	158	**19**
39	**0**	79	**0**	119	**20**	159	**17**
40	**0**	80	**0**	120	**17**	160	**16**
Average	**28.4**		**23.075**		**34.7**		**34.1**

Table 1: The number of correct answers given in the IQ experiment by participants according to treatments.

We use the nonparametric Mann-Whitney *U* test based on ranks to investigate whether the sample of correct answers comes from populations with the same median. We do a pairwise comparison by treatments. The test is appropriate because the distributions are clearly not normal. In Table 2 we report the test results. A number in the intersection of row and a column indicates, for the corresponding pair of treatments, the probability of getting at least an extreme absolute values of the test statistic as we observe, given that the two samples come from distributions with the same median.

	No payment (treatment 1)	10 cents (treatment 2)	1 NIS (treatment 3)
10 cents (treatment 2)	.0875	--	--
1 NIS (treatment 3)	.0687	.0004	--
3NIS (treatment 4)	.0708	.0006	.6964*

Table 2: Mann-Whitney *U* tests based on ranks with pairwise comparisons of medians of correct answers by treatment. (Prob. > |z|, where z is the test statistic). The * indicates that for that comparison we cannot reject (at a .9 level of significance) the hypothesis that the two samples comes from distribution with the same median.

The observation that the average number of correct answers decreased from 28.4 in the zero marginal payoff treatment to 23.075 when the marginal payoff was 10 cents seems particularly interesting. The difference between the two distributions is significant, at a .9 level of significance. When the marginal payoff was 1NIS and 3NIS, the average number of correct answers were 34.7 and 34.1 respectively. The difference between these two latter distributions is not significant, again at the .9 level of significance. Finally, the distributions in the high marginal payoff treatments (1NIS and 3NIS) are significantly higher than the distributions of the zero marginal payoff and the 10 cents marginal payoff.

Notice that none of the participants in the high payoff treatments, namely treatment 3 and 4, answered less than 16 correct answers. We may think that this is the minimal number of correct answers given by a participant who invest effort in solving the quiz. We may then wonder if the distribution of correct answer, conditional on investing effort (that is conditional on answering at least 16 questions) is different in the different treatments. Accordingly we performed this simple manipulation of the data: drop all observations in which participants answered less than 16 correct answers, and estimate the

distribution on the remaining data. The results of the statistical test are presented in Table 3.

	No payment (treatment 1)	10 cents (treatment 2)	1 NIS (treatment 3)
10 cents (treatment 2)	.2633*	--	--
1 NIS (treatment 3)	.4537*	.0680	--
3NIS (treatment 4)	.4633*	.1029*	.6964*

Table 3: Mann-Whitney *U* tests based on ranks with pairwise comparisons of medians of correct answers by treatment, after dropping all observations below 16. (Prob. > |z|, where z is the test statistic). The * indicates that for that comparison we cannot reject (at a significance level of .1) the hypothesis that the two samples comes from distribution with the same median.

The average number of correct answers in treatment 1 and 2 under this manipulation is 33.1 and 30.7 respectively. Of course, nothing is changed in treatments 3 and 4. From comparing the averages and from the results presented in table 3 we see that the difference between treatment 1 and 2 are much smaller and statistically insignificant. In fact, the only significant difference now is between treatment 2 and 3.

We conclude from this comparison that much of the difference observed between treatments is due to participants who decided not to put effort at all in the first two treatments. Note also that there is a substantial difference in the number of participants who answered less than 16 correct answers: 6 participants when the marginal payoff was zero and 11 when the marginal payoff was 10 cents.

A second interesting manipulation of the data is the comparison of the number of correct answers given by the 10 best participants in each treatment. The results of the statistical test are presented in Table 4.

	No payment (treatment 1)	10 cents (treatment 2)	1 NIS (treatment 3)
10 cents (treatment 2)	.1839*	--	--
1 NIS (treatment 3)	.1713*	.0084	--
3NIS (treatment 4)	.2536*	.0242	.4654*

Table 4: Mann-Whitney *U* tests based on ranks with pairwise comparisons of medians of correct answers by treatment, using only top 10 observations of each treatment. (Prob. > |z|, where z is the test statistic). * indicates that for that comparison we cannot reject (at a .9 level of significance) the hypothesis that the two samples comes from distribution with the same median.

The average number of correct answers is 43, 40.1, 45.5, and 45 in treatments 1, 2, 3, and 4 respectively. We conclude from this comparison that differences between treatments are observed in the top 10 observations as well. In particular, it is not the case that all the difference between treatments is due to participant who decided not to put effort at all (participants who answered less than 16 correct answers). Also within the top 10 participants, we observe a treatment effect.

The IQ experiment: the principals.

Now we report the answers of the subjects acting as principals in the IQ experiment. 46 out of the 53 (87%) subjects chose to pay 10 cents for every correct answer of the agent. That is, 87% of the participants preferred being matched with an agent who received a marginal payoff of 10 cents to being matched with an agent who received no marginal payoff at all. In fact, they are even willing to pay money for that purpose. This result indicates that the "economic assumption" is widely accepted even

when, it seems, it should not. People believe that the performance of individuals is monotonic in the money they receive: the higher the payoff the better the performance will be, irrespective of the amount. But we have seen that the increase in marginal payoff from zero to 10 cents actually lowered the number of correct answers given by participants significantly.

The donation experiment: the agents.

The amount of money collected by students is reported in Table 5.

10% (treatment 3a)	Obs.#	1% (treatment 2a)	Obs.#	No payment (treatment 1a)	Obs.#
0	61	**0**	31	**0**	1
0	62	**0**	32	**0**	2
20	63	**0**	33	**0**	3
20	64	**0**	34	**40**	4
30	65	**0**	35	**80**	5
50	66	**0**	36	**100**	6
100	67	**0**	37	**100**	7
100	68	**0**	38	**120**	8
120	69	**0**	39	**120**	9
140	70	**30**	40	**130**	10
150	71	**50**	41	**150**	11
150	72	**80**	42	**150**	12
150	73	**100**	43	**150**	13
150	74	**120**	44	**190**	14
180	75	**150**	45	**200**	15
200	76	**150**	46	**200**	16
200	77	**150**	47	**240**	17
240	78	**180**	48	**250**	18
250	79	**210**	49	**250**	19
250	80	**230**	50	**300**	20
290	81	**240**	51	**330**	21
290	82	**240**	52	**340**	22
350	83	**250**	53	**350**	23
380	84	**250**	54	**420**	24
400	85	**250**	55	**450**	25
410	86	**300**	56	**500**	26
460	87	**330**	57	**500**	27
500	88	**400**	58	**500**	28
500	89	**400**	59	**500**	29
500	90	**500**	60	**500**	30
219.33		**153.67**		**238.67**	**Average**

Table 5: The amount of money collected by students in the donation experiment according to treatments.

Again, we use the nonparametric Mann-Whitney *U* test based on ranks to investigate whether the sample of amounts of money donated comes from populations with the same median. The results of the test are reported in Table 6.

	No payment (treatment 1a)	1% (treatment 2a)
1% (treatment 2a)	.0977	--
10% (treatment 3a)	.7054*	.0515

Table 6: Mann-Whitney *U* tests based on ranks with pairwise comparisons of medians of amounts of money collected by treatment. (Prob. > |z|, where z is the test statistic). The * indicates that for that comparison we cannot reject (at a .9 level of significance) the hypothesis that the two samples comes from distribution with the same median.

We find special interest in the observation that when the payoff is increased from zero to 1% of the amount collected, the average collection decreased from 238.67 to 153.67. This difference is significant, at a .9 level of significance. When the payoff is increased to 10% of the amount collected, the average collection is 219.33. The amounts collected in this treatment is significantly higher than the amounts collected in the 1% treatment, but not significantly higher than the amounts collected when no payoff was given.

Like in the IQ experiment, we will perform some simple manipulations of the data. First, we exclude donations below 100 NIS from the analysis. The results of the test under these manipulations are presented in Table 7.

	No payment (treatment 1a)	1% (treatment 2a)
1% (treatment 2a)	.7887*	--
10% (treatment 3a)	.8173*	.6037*

Table 7: Mann-Whitney *U* tests based on ranks with pairwise comparisons of medians of amounts of money collected by treatment, excluding donation under 100NIS. (Prob. > |z|, where z is the test statistic). The * indicates that for that comparison we cannot reject (at a .9 level of significance) the hypothesis that the two samples comes from distribution with the same median.

The average amount collected after dropping all observations below 100 NIS is 281.6, 247.2, and 269.2 in treatments 1a, 1b, and 1c respectively. From table 7 we learn that the differences between treatments is no longer significant. Again, like in the IQ experiment, we conclude that much of the difference between treatments is due to people who invest very little effort or no effort at all.

To be consistent with the manipulations performed in the IQ test, we compare now the top 10 collections in each treatment. The statistical comparison is presented in Table 8.

	No payment (treatment 1a)	1% (treatment 2a)
1% (treatment 2a)	.0025	--
10% (treatment 3a)	.7593*	.0042

Table 8: Mann-Whitney *U* tests based on ranks with pairwise comparisons of medians of amounts of money collected by treatment, using only top 10 collections in each treatment. (Prob. > |z|, where z is the test statistic). The * indicates that for that comparison we cannot reject (at a .9 level of significance) the hypothesis that the two samples comes from distribution with the same median.

The average donation is 439, 316 and 408 in treatment 1a, 2a, and 3a respectively. As we see, there is a significant difference between the amounts collected when the payoff

was 1% and the amounts collected in the other two treatments. Like in the IQ experiment, this result implies that not all the difference between treatments is due to people who invest little effort. In particular, the highest collections of are also influenced by the experimental manipulation.

The donation experiment: the principals

Subjects acting as principals in the donation experiment behaved as the participants in the IQ experiment. 19 out of the 25 (76%) participants preferred to be matched with an agent who was paid 1% of the amount he collected. Recall that they had to pay 1% out of the 5% they received for this choice, while choosing to be matched with an agent that was not paid cost nothing. Like in the IQ experiment, we find this preferences especially interesting giving the results of the experiment with the agents.

5. CONCLUSIONS

The subjects in our experiments choose the level of effort they invest in answering the IQ quiz in one experiment, and in collecting donations in the other. The consequences of this action are the mild inconvenience, or possibly fun, in providing this effort; the signal that they receive on their own talent and skill in that particular task; and possibly, when it is offered, monetary rewards. Once the problem is formulated in these terms, the observed behavior of the participants in our experiments seems hard to explain. The addition of a small amount of money should not change the preference over outcomes, and if money is good, the effort should increase. (Note that this should be true even if subjects care about the opinion of the experimenter, because they should do so whether a monetary

reward is offered in addition or not.) In our experiments, on the contrary, the effort decreases when a small amount of money is added as reward.

Two different environments.

This experimental evidence suggests that the behavior of participants in our experiments is qualitatively different, depending on whether monetary incentives are used or not. Probably the most surprising result is the discontinuity of their behavior at the zero payment.

If we compare the situations in which some amount of money was offered, the performance improved as this amount increased. But the subjects in the experiments where no payment is offered, or more precisely no payment is even mentioned, perform better than the subjects in experiments where a small monetary compensation is offered.

We could not of course run an infinite number of experiments with smaller and smaller compensations: but it is natural to conjecture that the performance for amounts even smaller than the 10 cents would be even worse. This behavior would be consistent with the observed reduction in the level of performance as the amount paid was reduced. Note that the interesting simple model presented by Frey (1994) and Frey and Oberholzer-Gee (1997) does not predict this discontinuity. This of course is, strictly speaking, a technical consequence of the assumption of continuous preferences. A slightly different model, with a discontinuity at zero of the utility function would fit the data: but this would seem to be an artificial model.

It seems that the environment in which subjects were performing their activity (solving IQ quiz or collecting money for charitable institutions) was perceived in different

ways, depending on whether monetary compensation was offered or not. Whenever a monetary incentive was presented, the environment of the experiment was perceived as a monetary environment. When no monetary compensation was offered, or even mentioned, the environment was perceived as non-monetary. In the two cases we have two different curves of response of performance to monetary compensation.

Positive effects of the compensation in monetary environments.

Conditional on the perception that the environment was a monetary environment, however, the effect of monetary incentives is unambiguously positive: a larger monetary incentive improves the performance. In particular we could not find evidence of two positive levels of monetary compensation, where the performance with the higher compensation was lower. In fact it is surprising how effective the compensation is. For example in the IQ experiment moving from 10 cents to 1NIS seems enough to extract the maximum effort from the participants. We conclude that there are two distinct reaction curves: one describes the effort provided by people facing a monetary reward of different amount, and the other the (constant) effort provided by people that had no promise of monetary compensation. A precise estimate of these two curves is difficult, but we know that in the IQ experiments they cross at some amount between 10 cents and 1NIS.

Incentives and effort.

In our experiments we can measure the performance, not the effort that subjects put into the activity. Inspection of the data suggests the conjecture that subjects use a bang-bang procedure: they decide to put effort or not, and the incentives are effective

because they shift the subjects from a no-effort to an effort mode. To test this conjecture we analyzed the distribution of the performance above a given level. We tried to choose this level as the one achieved by a subject who is devoting at least a minimal effort to the task. The results are consistent with the bang-bang hypothesis both in the case of the IQ test experiments and in the donation experiments.

How small is a small amount? The fair wage hypothesis.

In our experiments the subjects who are paid 10 cents of a NIS for each right answer give a worse performance than those who are not paid at all. 10 cents may sound a very small compensation, almost "insulting''. While we believe that there is some truth in the intuition behind the word insulting, we also think that this is precisely what may be important in our results. Insulting compensations are not necessarily and always small compensations. The important fact is that the amount of money offered changes in a significant way the perception that the subjects have of "what the contract is about''. In particular it is not safe to assume that adding an incentive leaves the utility of the other incentives unchanged. A certain amount of monetary compensation may be perceived as too small when compared to the rest of the relevant factors, even if not too small in itself. We may think of real life situations where a non-trivial amount of money may sound disproportionately small compared to other factors. For instance, an increase in salary by two hundred dollars per month to a professor, as a compensation for a smaller office, may be worse than no compensation. This factor is likely to be more important when factors like health or reputation are at stake. So while in our experiments it is clear that ''too

small'' is somewhere between 10 cents and 1 NIS, the exact determination of this quantity in experimental or real-life situations is likely to be difficult and subtle.

How a principal perceives a contract.

In a second set of experiments we tried to test what is the perception that people have of the effect of monetary incentives on others. We did this by setting subjects as principals, paid according to the performance of agents. Agents were the subjects in the IQ experiments and the donation experiments that we have just described. Principals had to decide whether to offer a monetary compensation or not: this compensation was subtracted from their payoff.

We find that the subjects acting as principals make the wrong choice. They offer a compensation that reduces their payoff in two distinct ways. First, because they have to pay, and second because they are reducing the performance of their agents.

It is natural to explain the behavior of the principals with an argument similar to the one used to explain the behavior of the agents. Since they are offered the choice between a positive compensation and zero compensation, they perceive the environment as a monetary environment, and apply the "universal" rule that a larger compensation induces a better performance. Our results show that this rule is in effect justified, once the environment is recognized as a monetary environment. But the subjects fail to recognize the discontinuity in the behavior of their "agents". It is precisely this subtler phenomenon that seems not to be widely recognized.

Changes of preferences.

The results of our experiments might alternatively be explained in a simple, rational model of learning, as follows. Agents at the beginning of the experiments consider two states of nature as possible: one in which the activity will be performed in exchange for money, and the other in which it will not. Their preferences are defined over effort, money, and any other aspect of the outcome (as the opinion of the experimenters), and are different in the two states. Finally, upon receiving instructions subjects update their belief, and act accordingly. Clearly any amount of monetary compensation will be sufficient to inform them that the environment is monetary, and the true state is the first state.

This model explains the observed discontinuity. It is however very artificial, particularly in the second assumption: the utility of consequences is state dependent. In addition it has, potentially, a different major difficulty that we illustrate with a conjecture that we plan to test in future research.

Take the participants of the IQ quiz experiment, where 10 cents have been offered per each question answered correctly, and in a second experiment ask them to perform the same task, for no compensation. We conjecture that the performance will be worse than the one of subjects who have been offered zero compensation from the start, and of course worse than the performance of the same participants in the first experiment. This behavior would not be consistent with the model we have just discussed. Rather, it seems to suggest that as long as no monetary compensation is announced, subjects are unaware of this possibility. We conjecture that the perception of subjects of the contract changes during the experiment. As we mentioned earlier, it would be interesting to determine if the perception of the task or rather the perception of the environment changes. This seems possible to test experimentally, by performing the same experiment we have just described,

but changing the task. More complex, and more interesting, is to determine the change in performance induced by a protracted offer of a money reward, but, as we said, this is topic for future research.

Appendix: Instruction for the IQ experiment

Introduction

The instructions are simple, and if you follow them carefully you may earn a considerable amount of money. The experiment would take about 45 minutes.

In the experiment you are asked to answer a quiz of 50 problems taken out of a psychometric test used to scan candidates to the university. This is a sort of an IQ test.

You will be paid 60NIS for showing up to the experiment. {*The following sentence was not included in treatment 1*: ‘’ In addition, you will be paid 0.1NIS’’ (in treatment 2, 1NIS in treatment 3 and 2NIS in treatment 4) for every correct answer you will give.}

The money will be paid to you, privately and in cash, at the end of the experiment.

Do you have any question?

REFERENCES

Akerlof, G. A. (1982) “Labor contracts as partial gift exchange,” *Quarterly Journal of Economics*, 543-569.

Akerlof, G. A. and J. L. Yellen (1988) “Fairness and unemployment,”*American Economic Review*, 44-49.

Akerlof, G. A. and J. L. Yellen (1990) “The fair wage-effort hypothesis and unemployment,” *Quarterly Journal of Economics*, 255-283.

Allport, G. W., (1937), *Personality: A Psychological Interpretation*, New York, Holt.

Arrow, K. J., (1972) "Gifts and Exchanges," *Philosophy and Public Affairs*, 1,2, 343-362.

Bem, D. J., (1965) "An Experimental Analysis of Self-Persuasion," *Journal of Experimental and Social Psychology*, 1, 199-218.

Bem, D. J., (1967) "Self-Perception: An alternative interpretation of Cognitive dissonance phenomena," *Psychological Review*, 74, 183-200.

Deci, Edward, (1971) "Effects of externally mediated rewards on intrinsic motivation," *Journal of Personality and Social Psychology*, 18, 105-115.

Deci, E. and Cascio, E.L. (1972) "Changes in Intrinsic Motivation as a function of negative feedback and threats," Paper presented at the meeting of the Eastern Psychological Association, Washington, April 1972.

Deci, E., Cascio, E.L. and Krusell, (1973) "Sex differences, verbal reinforcement and intrinsic motivation," Paper resented at the meeting of the Eastern Psychological Association, Washington, May 1973.

Deci, Edward, (1975) *Intrinsic Motivation*, Plenum Press, New York and London.

Festinger, L., (1957), *A Theory of Cognitive Dissonance*, Evanston, Illinois: Row, Peterson.

Frey, B.S. (1994) "How Intrinsic Motivation is Crowded in and Out," *Rationality and Society*, 6, 3, 334-352.

Frey B. S., Oberholzer F. and Eichenberger R., (1996) "The old lady visits your backyard: a tale of morals and markets," *Journal of Political Economy*, 104, 6, 1297-1313.

Frey B. S. and F. Oberholzer-Gee, (1997) "The Cost of Price Incentives: An Empirical Analysis of Motivation Crowding-Out," *American Economic Review*, 87, 4, 746-755.

Greene, D., Sternberg, B. and Lepper, M. R. (1976) "Overjustification in a token economy," *Journal of Personality and Social Psychology*, 34, 1219-1234.

Heider F., (1958), *The Psychology of Interpersonal Relations*, Wiley,

Kelley, H. H., (1967) "Attribution Theory in Social Psychology," *Nebraska Symposium on Motivation*, 15, 192-238.

Kelley, H. H., (1971), *Attribution in Social Interaction*, New York: General Learning Press, Module Series.

Kruglansky, A. Freedman, and Zeevi, (1971) "The effects of extrinsic incentives on some qualitative aspects of task performance," *Journal of Personality*, 39, 606-617

Kruglansky, A. W. Alon, S. and Lewis, T. (1972) "Retrospective misattribution and task enjoyment," *Journal of Experimental Social Psychology*, 8, 493-501.

Lepper, M.R., Greene, D. and Nisbett, R.E., (1973) "Undermining children's intrinsic interest with extrinsic rewards: A test of the "overjustification" hypothesis," *Journal of Personality and Social Psychology*, 28, 129-137.

Lepper, Mark R. and Greeene, David, (1978), *The Hidden Costs of Reward: New Perspectives in the Psychology of Human Motivation*, Lawrence Elbaum Associates, Publishers; John Wiley and Sons.

Reiss S. and Sushinsky, L. W. (1975a) "Overjustification, competing responses and the acquisition of intrinsic interest," *Journal of Personality and Social Psychology*, 31, 1116-1125.

Reiss S. and Sushinsky, L. W. (1975b) "Comment: Undermining extrinsic interest," *American Psychologist*, 31, 782-783.

Smith V., (1982) "Microeconomic Systems as an Experimental Science", *American Economic Review*, December 1982, 72, 923-955.

Titmuss, R. M., (1970) *The Gift Relationship*, London:, Allen and Unwin.

Comparison of the top 20 subjects in each treatment

	No payment (treatment 1)	10 cents (treatment 2)	1 NIS (treatment 3)
10 cents (treatment 2)	.0381	--	--
1 NIS (treatment 3)	.0146	.0003	--
3NIS (treatment 4)	.0339	.0008	.4222*

Comparison of the worse 20 subjects in each treatment

	No payment (treatment 1)	10 cents (treatment 2)	1 NIS (treatment 3)
10 cents (treatment 2)	.0299	--	--
1 NIS (treatment 3)	.0120	.0000	--
3NIS (treatment 4)	.0270	.0001	.7552*